

# Out-of-plane heat transfer in van der Waals stacks: electron-hyperbolic phonon coupling

K. J. Tielrooij,<sup>1,\*</sup> N. C. H. Hesp,<sup>1</sup> A. Principi,<sup>2</sup> M. Lundeberg,<sup>1</sup> E. A. A. Pogna,<sup>3</sup> L. Banszerus,<sup>4</sup> Z. Mics,<sup>5</sup> M. Massicotte,<sup>1</sup> P. Schmidt,<sup>1</sup> D. Davydovskaya,<sup>1</sup> K. Watanabe,<sup>6</sup> T. Taniguchi,<sup>6</sup> M. Bonn,<sup>5</sup> D. Turchinovich,<sup>5</sup> C. Stampfer,<sup>4</sup> G. Cerullo,<sup>3</sup> M. Polini,<sup>7</sup> and F. H. L. Koppens<sup>1,8,\*</sup>

<sup>1</sup>*ICFO - Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Castelldefels (Barcelona) 08860, Spain*

<sup>2</sup>*Radboud University, Institute for Molecules and Materials, NL-6525 AJ Nijmegen, the Netherlands*

<sup>3</sup>*IFN-CNR, Dipartimento di Fisica, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy*

<sup>4</sup>*JARA-FIT and 2nd Institute of Physics, RWTH Aachen University, 52074 Aachen, Germany*

<sup>5</sup>*Max Planck Institute for Polymer Research, Ackermannweg 10, Mainz 55128, Germany*

<sup>6</sup>*National Institute for Material Science, 1-1 Namiki, Tsukuba 305-0044, Japan*

<sup>7</sup>*Istituto Italiano di Tecnologia, Graphene Labs, Via Morego 30, I-16163 Genova, Italy*

<sup>8</sup>*ICREA - Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain*

Van der Waals heterostructures have emerged as promising building blocks that offer access to new physics, novel device functionalities, and superior electrical and optoelectronic properties [1–6]. Applications such as thermal management, photodetection, light emission, data communication, high-speed electronics and light harvesting [7–14] require a thorough understanding of (nanoscale) heat flow. Here, using time-resolved photocurrent measurements we identify a surprisingly efficient out-of-plane energy transfer channel, where charge carriers in graphene couple to hyperbolic phonon polaritons [15–17] in the encapsulating layered material. This hyperbolic cooling is particularly efficient – giving picosecond cooling times – for hexagonal BN, where the high-momentum hyperbolic phonon polaritons enable efficient near-field energy transfer. We study this heat transfer mechanism through distinct control knobs, such as carrier density and lattice temperature, and find excellent agreement with theory without any adjustable parameters. These insights will lead to the ability to control heat flow in van der Waals heterostructures.

Owing to its large in-plane thermal conductivity, graphene has been suggested as preferred material for thermal management of nanoscale devices [7]. At the same time, graphene is well-known for its ability to convert incident light into electrical heat, i.e. hot electrons that can be used to generate photocurrent, with applications in photodetection, data communication and light harvesting [9, 18, 19]. Understanding – and ultimately controlling – heat flow in graphene–van der Waals heterostructures is therefore of paramount importance. For example, a short cooling time of graphene hot carriers is advantageous for thermal management and for high switching rates of photodetectors for data communication, whereas a long cooling time is favorable for photodetection sensitivity [9, 18, 19]. Of particular relevance are heterostructure devices that contain high-quality graphene encapsulated by layered materials, such as hexagonal BN (hBN) and MoS<sub>2</sub>, which have the potential to crucially improve the performance of a myriad of electronic and optoelectronic devices [2].

A large number of cooling pathways for graphene carriers have been proposed, involving among others strongly coupled optical phonons [20], graphene intrinsic

acoustic phonons [21–25], substrate phonons [26] and plasmons [27]. Here, using several experimental approaches, we show that cooling in low-disorder graphene (by encapsulation in hBN) is governed by out-of-plane coupling of graphene electrons to special polar phonon modes that occur in layered materials: hyperbolic phonon polaritons, where  $\epsilon_{xx}\epsilon_{zz} < 0$ , with  $\epsilon_{xx}$  and  $\epsilon_{zz}$  the permittivity parallel and perpendicular to the two-dimensional material plane. Owing to this property, these materials carry deep sub-wavelength, ray-like optical phonon polaritons. For hBN, within the two Reststrahlen bands a large number of hyperbolic phonon modes exist with high momenta, far outside the light cone. The most notable modes occur at an energy of  $\sim 100$  meV and  $\sim 180$  meV [15], such that energy overlap with the graphene hot-carrier distribution is substantial. The unusual hyperbolic character gives rise to a very high density of optical states, and thus large thermal energy densities [15, 16], thereby providing a potentially efficient cooling pathway for hot carriers in graphene. By near-field coupling between graphene and hBN, efficient energy transfer from hot carriers to hyperbolic phonon polaritons is possible [28]. We will show below that the measured carrier dynamics of hBN-encapsulated graphene can indeed be fully explained by this hyperbolic cooling process, as illustrated in Fig. 1a.

\*Electronic address: Correspondence: klaas-jan.tielrooij@icfo.eu, frank.koppens@icfo.eu

We use ultrafast time-resolved photocurrent mea-

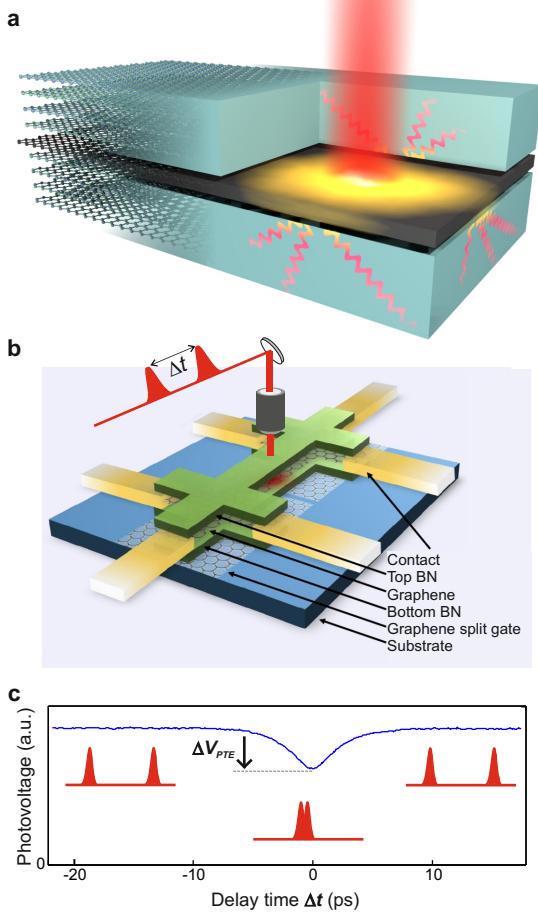


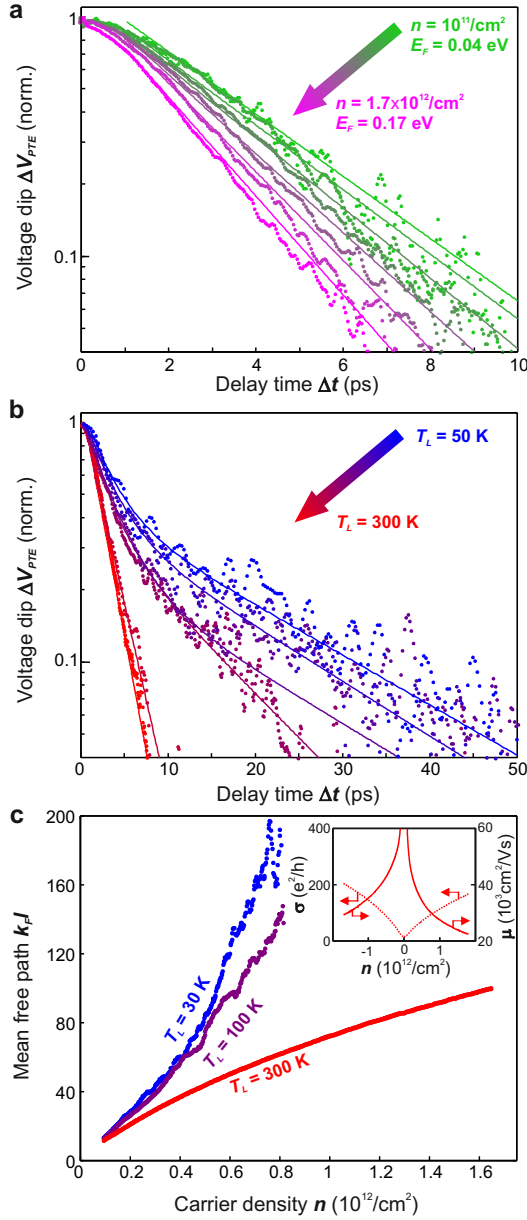
FIG. 1: **Hot carrier cooling in hBN-encapsulated graphene.** **a)** Schematic representation of out-of-plane heat transfer in hBN-encapsulated graphene: highly efficient near-field emission from graphene hot carriers into directional hyperbolic phonon polaritons of the encapsulating material. **b)** Schematic drawing of the hBN-encapsulated graphene device with graphene split gates. By applying a different voltage to the two split gates we create a *pn*-junction in the encapsulated graphene layer, which we illuminate with two ultrafast pulses that arrive with a variable time delay  $\Delta t$ . **c)** The resultant photo-thermoelectric photovoltage measured between two contacts shows a dip when the two ultrafast pulses overlap. This is due to the nonlinear relation between hot carrier temperature and incident light intensity. The dynamics of the photovoltage dip  $\Delta V_{PTE}$  as a function of time delay  $\Delta t$  therefore correspond to the cooling dynamics of the hot carriers at the *pn*-junction.

measurements on an hBN-encapsulated, Hall-bar shaped, exfoliated graphene device (see Fig. 1b). The device contains CVD graphene split gates underneath the bottom hBN, allowing the generation of a *pn*-junction in the middle of the device. With 800 nm light incident on the *pn*-junction, a photovoltage is generated by the photo-thermoelectric (PTE) effect [18, 19] (see the photocurrent and reflection data in Fig. A1 and the characteristic PTE sixfold pattern in Fig. A2). By varying the delay  $\Delta t$  between two sub-picosecond pulses

we extract the carrier dynamics from the photovoltage signal  $\Delta V_{PTE}(\Delta t)$ . Since the photo-thermoelectric voltage scales with the light-induced increase in carrier temperature, the decay dynamics of  $\Delta V_{PTE}(\Delta t)$  corresponds to the cooling dynamics of the hot electron system (See Fig. 1c and Fig. A3). In the case of exponential decay dynamics, which we observe above a lattice temperature of  $\sim 200$  K, we extract an experimental cooling time scale  $\tau_{\text{exp}}$  by describing the decay dynamics with  $\Delta V_{PTE}(\Delta t) \propto e^{-\Delta t/\tau_{\text{exp}}}$ .

To study hot-carrier cooling we first examine  $\Delta V_{PTE}(\Delta t)$  while varying graphene's most characteristic parameter, the carrier density  $n$ . In particular, we apply a gate voltage  $V_L = +V - V_D$  to the left split gate and  $V_R = -V - V_D$  to the right gate, such that there is always either a *pn*-junction or an *np*-junction, with equal electron and hole densities in the two graphene regions ( $V_D$  is the gate voltage that corresponds to the Dirac point). The incident laser fluence is typically  $5 - 40 \mu\text{J}/\text{cm}^2$ . The data show that cooling becomes faster upon increasing the carrier density (see Fig. 2a). We also vary the lattice temperature  $T_L$  and observe faster decay for increasing lattice temperature (see Fig. 2b, taken at  $n \approx 10^{12}/\text{cm}^2$ ). At low temperatures (below  $\sim 200$  K), cooling is clearly non-exponential, whereas at room temperature we observe exponential decay of the photovoltage dip, with a timescale of  $\tau_{\text{exp}} \approx 2.5$  ps (for  $n = 1.7 \times 10^{12}/\text{cm}^2$ ). We independently verify this cooling time using two alternative measurement techniques that are each sensitive to electron cooling dynamics in a different way (see Fig. A4). Using ultrafast optical pump – optical probe spectroscopy, which probes interband transitions [29], we find  $\tau_{\text{exp}} = 2.55$  ps for the decay of the absorption photobleaching; using optical pump – terahertz probe spectroscopy, which probes intraband transitions [12], we obtain  $\tau_{\text{exp}} = 2.2$  ps for the decay of the photoconductivity. For this we used very similar excitation conditions (an incident pulse fluence of  $8 - 20 \mu\text{J}/\text{cm}^2$ ) as in the photocurrent measurements, and two separate devices with large-area, high-quality hBN-encapsulated CVD graphene as in Ref. [30]. Thus, all three techniques consistently yield a very similar cooling time for hBN-encapsulated graphene. Interestingly, the observed cooling timescale is almost identical to the cooling time of  $\sim 2$  ps for hot carriers in  $\text{SiO}_2$ -supported graphene, where supercollision cooling dominates [23–25]. Therefore, we will first assess to what extent our data are compatible with supercollisions as the main cooling pathway.

The supercollision cooling mechanism relies on the presence of short-range scatterers (long-range scattering mechanisms give infinite cooling time) [22] and gives a cooling time that scales with  $\frac{k_F \ell}{D^2}$ , where  $k_F$  is the Fermi momentum,  $\ell$  the mean free path (limited by short-range scattering), and  $D$  the electron-phonon deformation potential [23]. We performed four-probe measurements



**FIG. 2: The effect of doping and lattice temperature.** The dynamics of the photovoltage dip  $\Delta V_{\text{PTE}}$  show that hot-carrier cooling becomes faster upon increasing the carrier density (a) and upon increasing the lattice temperature (b). Panel a is measured at a lattice temperature of  $T_L = 300$  K and panel b at a carrier density of  $n \approx 1.7 \times 10^{12}/\text{cm}^2$ . The solid lines are bi-exponential guides to the eye. c) From four-probe transport measurements, we extract the dimensionless mean free path parameter  $k_F\ell$  as a function of carrier density for three different lattice temperatures. These data show that carrier scattering in our hBN-encapsulated graphene is much less efficient than in  $\text{SiO}_2$ -supported graphene where  $k_F\ell$  is typically 10 or less, and where disorder-assisted supercollisions constitute the dominant cooling mechanism [22–24]. The inset shows the conductivity and mobility vs. gate-induced carrier density.

(see Fig. 2c) on our device and find  $k_F\ell = 80$ –100 at

a carrier density of  $n = 1.7 \times 10^{12}/\text{cm}^2$ , corresponding to a mobility of 25,000–30,000  $\text{cm}^2/\text{Vs}$  and momentum scattering time of 340–440 fs. This is indeed much higher than  $\text{SiO}_2$ -supported devices with typically  $k_F\ell < 10$ , mobility  $< 5,000$   $\text{cm}^2/\text{Vs}$  and momentum scattering time  $< 100$  fs (see e.g. Ref. [23]). The value of  $D$  is reasonably well established, with transport measurements on ultraclean, hBN-encapsulated devices giving  $\sim 18$ –20 eV, assuming phonon-limited momentum scattering [2, 6]. In accordance, it was found that the cooling dynamics of  $\text{SiO}_2$ -supported graphene are consistent with  $D = 12$ –18 eV, assuming disorder-assisted cooling [23]. From our transport measurements we find  $D < 35$  eV, assuming phonon-limited momentum scattering (see Methods). This is an *upper* bound, because long- and short-range scattering mechanisms also limit the momentum scattering time. For our  $k_F\ell$  and with  $D = 12$ –35 eV, we obtain a cooling time of roughly 4–20 ps if disorder-assisted cooling would dominate. Indeed, we find that in order to reproduce the observed cooling dynamics, the deformation potential would have to be an unrealistically high  $D > 65$  eV (see Fig. A5 and Methods). This value is a *lower* bound, because the analysis assumes that all scatterers that lead to the mean free path also contribute to supercollision scattering, whereas actually only a fraction – the short-range scatterers – contribute [22]. Since transport measurements indicate  $D < 35$  eV and cooling dynamics indicate  $D > 65$  eV, we conclude that supercollision is very likely *not* the dominant cooling mechanism in hBN-encapsulated graphene. This is corroborated by the observed trend that cooling becomes faster with increasing carrier density – exactly opposite to the trend that is measured for supercollision-dominated devices where faster cooling is observed around the Dirac point [23, 25]. Finally, a recent noise thermometry study on hBN-encapsulated graphene also led to the conclusion that supercollision cooling cannot explain the observed thermal conductance [31].

Based on these considerations, we propose an out-of-plane cooling mechanism where hot carriers in graphene lose their energy to remote polar phonons in the encapsulating layered material. The energy transfer can be understood from fluctuation electrodynamics, where any process that dissipates energy in the form of heat, has a reverse process that is driven by thermal fluctuations and thus becomes stronger at higher temperatures [32]. An example of such paired processes is light absorption and blackbody radiation, which means that graphene emits thermal noise due to the dissipative real part (indicated by  $\mathcal{R}$ ) of the frequency- and momentum-dependent optical sheet conductivity  $\sigma(\omega, k)$ . This thermal noise is efficiently absorbed by hBN, hBN being a lossy-polarizable

material, leading to an energy transfer rate [28]:

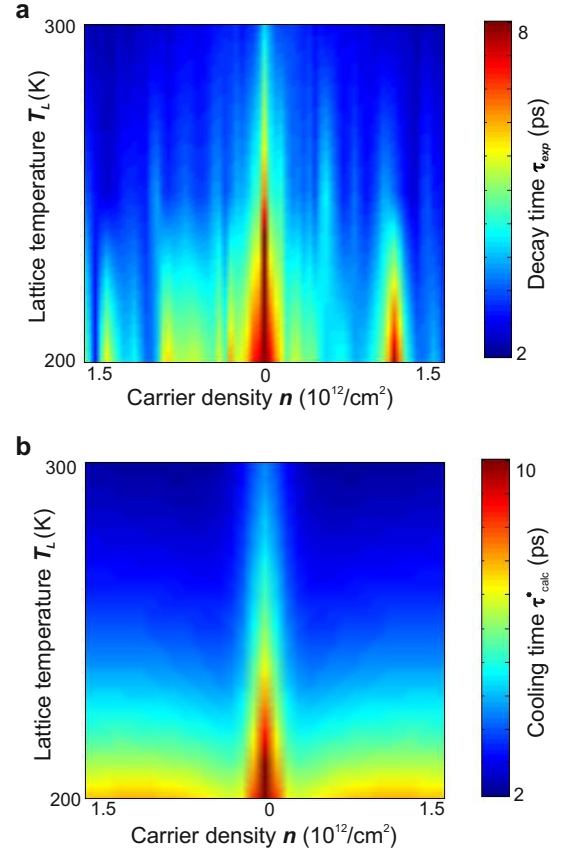
$$\mathcal{Q} = \iiint_{-\infty}^{\infty} \frac{d\omega dk_x dk_y}{(2\pi)^3} [n_B(\omega, T_e) - n_B(\omega, T_L)] M(\omega, k) \quad , \quad (1)$$

where  $n_B(\omega, T) = \frac{\hbar|\omega|}{e^{\hbar|\omega|/k_B T} - 1}$  and  $M(\omega, k) = 4 \frac{\mathcal{R}\{Y(\omega, k)\} \mathcal{R}\{\sigma(\omega, k)\}}{|Y(\omega, k) + \sigma(\omega, k)|^2}$  ( $k_B$  is Boltzmann's constant). Heat transfer from hot graphene electrons to hBN hyperbolic phonon polaritons is thus governed by the Bose factor  $[n_B(\omega, T_e) - n_B(\omega, T_L)]$ , which describes the energy disequilibrium between hot graphene carriers and the cold hBN phonon system, and the impedance matching function  $M(\omega, k)$ , which is nonzero when the surface admittance  $Y(\omega, k)$  has a nonzero real part. We calculate  $Y(\omega, k)$  as in Ref. [33] and find that nearby the hyperbolic hBN phonon frequencies  $Y(\omega, k)$  is real over a large  $k$ -space area and relatively wide frequency band. Since  $\sigma(\omega, k)$ , calculated using the Random Phase Approximation [34], also has a significant real part, this leads to an impedance matching function  $M$  approaching unity. Due to this near-field coupling to hyperbolic modes, the heat conductivity exceeds Planck's law for blackbody radiation by orders of magnitude (see Fig. A6). The reason for this is that in vacuum, the  $k$ -space for blackbody radiation is limited to  $k < \omega/c$  (with  $c$  the speed of light), whereas this restriction is lifted in the near-field interaction with hBN hyperbolic phonon polaritons. This super-Planckian coupling to hyperbolic hBN phonons thus provides a highly efficient cooling channel for hot carriers in graphene. Cooling to hyperbolic modes also occurs in materials such as MoS<sub>2</sub>, although there it is not as efficient as for hBN (see Fig. A6.)

To compare our hyperbolic cooling theory with the experimental data, we examine the calculated energy transfer rate  $\mathcal{Q}$ . Here, solving Eq. (1) gives cooling dynamics with a cooling time scale

$$\tau_{\text{calc}}(T_e, T_L) = C_n \frac{T_e - T_L}{\mathcal{Q}} \quad , \quad (2)$$

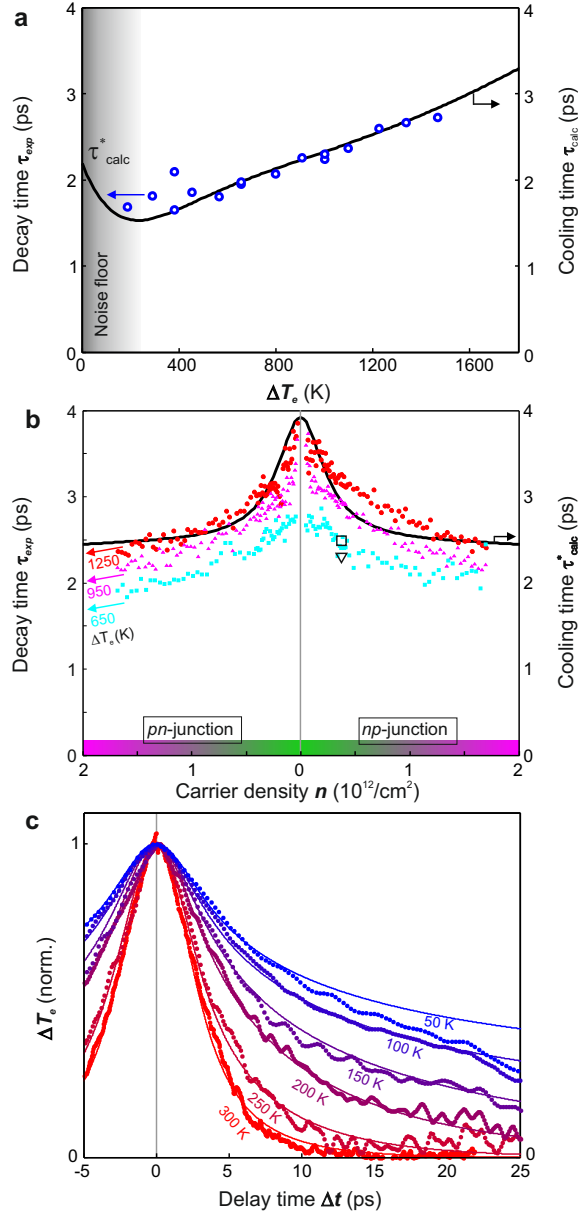
where  $C_n$  is the electronic heat capacity of graphene at constant carrier density  $n$  [28]. In the limit of weak heating, where  $T_e$  approaches  $T_L$ , we obtain exponential decay with the near-equilibrium timescale  $\tau_{\text{calc}}(T_e \rightarrow T_L) = \tau_{\text{calc}}^* = C_n \left( \frac{\partial \mathcal{Q}}{\partial T_e} \Big|_{T_e=T_L} \right)^{-1}$ , where  $\frac{\partial \mathcal{Q}}{\partial T_e} \Big|_{T_e=T_L}$  is the interfacial heat conductivity  $\Gamma$ . We compare the calculated near-equilibrium cooling time  $\tau_{\text{calc}}^*$  with the measured exponential decay time  $\tau_{\text{exp}}$  (see Fig. 3), although strictly speaking these experimental timescales correspond to the strong heating regime, where  $T_e \gg T_L$ . The reason for this is that our technique is not sensitive enough at low incident powers, which means that experimentally we cannot directly access the near-equilibrium cooling time  $\tau_{\text{calc}}^*$ . Nevertheless, we find that our hyperbolic hBN cooling model semi-quantitatively reproduces the experimen-



**FIG. 3: Qualitative comparison with hyperbolic hBN cooling.** Comparison between the experimental decay time  $\tau_{\text{exp}}$  extracted from the dynamics of the photovoltage dip (a) and the predicted near-equilibrium cooling time  $\tau_{\text{calc}}^*$  for super-Planckian cooling to hyperbolic hBN phonons (b). Both in experiment and theory, we vary the carrier density and lattice temperature.

tally observed trends for the entire range of investigated carrier densities (up to  $1.6 \times 10^{12}/\text{cm}^2$ ) and lattice temperatures (200 – 300 K). In particular, cooling slows down for lower lattice temperatures, which we attribute mainly to the smaller Bose factor in Eq. (1) and thus smaller energy transfer rate  $\mathcal{Q}$ . The calculations also reproduce the striking observation of a longer cooling time around the Dirac point. This is the result of the energy transfer rate  $\mathcal{Q}$  smoothly decreasing towards zero carrier density, while the electronic heat capacity  $C_n$  flattens around the Dirac point towards its neutral graphene value (see Fig. A6). At higher carrier densities ( $n > 10^{12}/\text{cm}^2$ ) the increasing energy transfer rate is compensated by the increasing heat capacity, leading to a fairly weak dependence of cooling time on carrier density.





**FIG. 4: Quantitative comparison with hyperbolic hBN cooling.** **a)** Comparison between measured decay time  $\tau_{\text{exp}}$  (blue symbols and left vertical axis) and calculated cooling time  $\tau_{\text{calc}}$  (solid black line and right vertical axis) as a function of electron temperature *without any adjustable parameters*. **b)** The measured decay time  $\tau_{\text{exp}}$  (symbols and left vertical axis) at room temperature as a function of carrier density for three different laser powers, compared with the near-equilibrium cooling time  $\tau_{\text{calc}}$  (solid line and right vertical axis) according to super-Planckian cooling to hyperbolic hBN phonons *without any adjustable parameters*. The open square (triangle) shows the obtained decay time from time-resolved optical (terahertz) spectroscopy for  $\Delta T_e = 600 \pm 200$  K. **c)** Comparison of the complete cooling dynamics as measured (data points) and calculated (solid line) for  $n = 0.06 \times 10^{12}/\text{cm}^2$  for varying lattice temperature. The calculated dynamics are convoluted with a Gaussian function representing the experimental time resolution. We find an initial temperature increase of  $\Delta T = 300$ – $500$  K for all temperatures, except 1250 K for  $T_L = 300$  K. The lower  $\Delta T$  is related to the larger effective photocurrent spot size.

To make a more quantitative comparison, we take into account that the measurements are typically done in the strong heating regime, where  $T_e \gg T_L$ . We first measure the cooling dynamics for increasing laser power and estimate the electron temperature that corresponds to each power from the characteristic power-dependent photoresponse (see Fig. A7 and Methods), thus obtaining the exponential decay time  $\tau_{\text{exp}}$  vs.  $T_e$  (see Fig. 4a). In the experimentally accessible regime ( $\Delta T_e = T_e - T_L > 200$  K) the decay time  $\tau_{\text{exp}}$  increases with increasing  $T_e$  (measured for  $n = 1.2 \times 10^{12}/\text{cm}^2$ ). We compare this with the calculated cooling time  $\tau_{\text{calc}}$ , which describes the “instantaneous” cooling time at a certain  $T_e$ , and find quantitative agreement *without any adjustable parameters*. The reason for the increasing cooling time with increasing  $T_e$  can be seen from Eq. (2) and by noting that  $Q$  scales with  $(T_e - T_L)$  for large  $T_e$ . This means that the temperature scaling is determined by the electronic heat capacity  $C_n$ , which increases with increasing  $T_e$ . The agreement between experiment and calculation prompts us to make a more quantitative comparison for varying carrier density  $n$ . Figure 4b shows  $\tau_{\text{exp}}$  for three different laser powers ( $P = 21$ , 47 and 94  $\mu\text{W}$ ), corresponding to three different initial hot electron temperatures ( $\Delta T_e = 650$ , 950 and 1250 K), together with the calculated near-equilibrium cooling time  $\tau_{\text{calc}}^*$ . The comparison between experimental results in the strong heating regime ( $\tau_{\text{exp}}$ ) and theoretical results close to equilibrium ( $\tau_{\text{calc}}^*$ ) is justified in Fig. 4a, where we show that the cooling time at  $T_e \sim 1000$  K is very similar to  $\tau_{\text{calc}}^*$ . Again we find agreement between experiment and theory *without any adjustable parameters*.

Finally, we compare the calculated time-domain cooling dynamics in the strong heating regime directly with the measured photocurrent dynamics. The calculated cooling dynamics describe cooling from an initial  $T_e$  down to  $T_L$ , as constructed from the electron temperature-dependent cooling time  $\tau_{\text{calc}}$  in Eq. (2). We compare these calculated cooling dynamics directly with the measured photovoltage dynamics  $\Delta V_{\text{PTE}}(\Delta t)$  for six different lattice temperatures, at a low carrier density of  $n = 0.06 \times 10^{12}/\text{cm}^2$  (see Fig. 4c). Using the initial temperature as a fit parameter, we find that the hyperbolic cooling model can describe the experimental data very well. At room temperature, the initial temperature is similar to the one we calculate (see Fig. A7), whereas we find a lower initial temperature at lower lattice temperatures. This is due to the increased mechanical vibrations of the sample and lateral heat diffusion out of the laser spot at lower  $T_L$ , which both lead to a larger photocurrent spot size and thus a lower effective initial  $T_e$  (see Fig. A8). For a lattice temperature of 50 K, we furthermore see an increasing discrepancy for delay time  $> 15$  ps, which most likely indicates that hBN-phonons alone do not account for the complete cooling dynamics. Similarly, we find that

at a higher carrier density of  $n = 1.7 \times 10^{12}/\text{cm}^2$  the hyperbolic cooling model starts deviating from the experimental data already for  $T_L < 200$  K (see Fig. A9). Most likely, at sufficiently low lattice temperature and sufficiently high carrier density, momentum-conserving cooling to acoustic graphene phonons becomes the dominant cooling channel [21, 35, 36].

*Conclusion* — This study has addressed the issue of out-of-plane heat transfer in the highly promising device architecture of hBN-encapsulated graphene. Combining experimental results with microscopic theoretical calculations, we have been able to show that the dominant cooling channel is the one in which heat transfer from hot carriers to the hBN polar substrate occurs via near-field coupling to hyperbolic phonon modes. This remarkably efficient mechanism explains the surprising observation of a lower-than-expected cooling time in clean hBN-encapsulated graphene. We predict that such devices will exhibit significantly slower cooling by using hBN encapsulation with very thin hBN or alternative layered dielectrics, such as  $\text{MoS}_2$ , which is highly relevant for photodetection applications. Furthermore, the near-field coupling we studied in this work between hot graphene carriers and hyperbolic hBN phonons may pave the way to novel approaches in fields such as nanophotonics, ultrahigh resolution microscopy

and nanoscale thermal management.

*Acknowledgements* — This work was supported by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 696656 Graphene Flagship, Fondazione Istituto Italiano di Tecnologia, the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0522), Fundacio Cellex Barcelona, the Mineco grants Ramon y Cajal (RYC-2012-12281), Plan Nacional (FIS2013-47161-P), and the Government of Catalonia through the SGR grant (2014-SGR-1535), the ERC StG CarbonLight (307806). K.J.T. acknowledges support through the Mineco Young Investigator Grant (FIS2014-59639-JIN). A.P. acknowledges support from the ERC Advanced Grant 338957 FEMTO/NANO and from the NWO via the Spinoza Prize. M.M. thanks the Natural Sciences and Engineering Research Council of Canada (PGSD3-426325-2012). D.T. acknowledges financial support from European Union Marie Curie Program (Career Integration Grant No. 334324 LIGHTER) and Max Planck Society. K.W. and T.T. acknowledge support from the Elemental Strategy Initiative conducted by the MEXT, Japan and JSPS KAKENHI Grant Numbers JP26248061, JP15K21722 and JP25106006.

- 
- [1] Geim, A.K. & Grigorieva, I.V. Van der Waals heterostructures *Nature* **499**, 419-425 (2013)
  - [2] Dean, C.R. et al. Boron nitride substrates for high-quality graphene electronics. *Nature Nanotech* **5**, 722 (2010)
  - [3] Mayorov, A.S. et al. Micrometer-scale ballistic transport in encapsulated graphene at room temperature. *Nano Lett.* **11**, 2396 (2011)
  - [4] Britnell, L et al. Electron tunneling through ultrathin boron nitride crystalline barriers. *Nano Lett.* **12**, 1707 (2012)
  - [5] Dean, C.R. et al. Hofstadter’s butterfly and the fractal quantum Hall effect in moiré superlattices. *Nature* **497**, 598 (2013)
  - [6] Wang, L. et al. One-Dimensional Electrical Contact to a Two-Dimensional Material. *Science* **342**, 614-617 (2013)
  - [7] Balandin, A.A. et al. Superior Thermal Conductivity of Single-Layer Graphene. *Nano Lett.* **8**, 902–907 (2008)
  - [8] Britnell, L. et al. Strong Light-Matter Interactions in Heterostructures of Atomically Thin Films. *Science* **340** 1311–1314 (2013)
  - [9] Koppens, F.J.L. et al. Photodetectors based on graphene, other two-dimensional materials and hybrid systems. *Nature Nanotech.* **9**, 780793 (2014)
  - [10] Lopez-Sanchez, O. et al. Light Generation and Harvesting in a van der Waals Heterostructure. *ACS Nano* **8**, 3042 (2014)
  - [11] Bonaccorso, F. et al. Graphene, related two-dimensional crystals, and hybrid systems for energy conversion and storage. *Science* **347**, 1246501 (2015)
  - [12] Mics, Z et al. Thermodynamic picture of ultrafast charge transport in graphene. *Nature Comm.* **6**, 7655 (2015)
  - [13] Kim, Y.D. et al. Bright visible light emission from graphene. *Nature Nanotech.* **10**, 676-681 (2015)
  - [14] Massicotte, M. et al. Picosecond photoresponse in van der Waals heterostructures. *Nature Nanotech.* **11**, 42–46 (2016)
  - [15] Caldwell, J.D. et al. Sub-diffractive volume-confined polaritons in the natural hyperbolic material hexagonal boron nitride. *Nature Comm.* **5**, 5221 (2014)
  - [16] Dai, S. et al. Tunable phonon polaritons in atomically thin van der Waals crystals of boron nitride. *Science* **343** 1125 (2014)
  - [17] Basov, D.N., Fogler, M.M. & Garcia de Abajo, F.J. Polaritons in van der Waals materials. *Science* **354**, 1992 (2016)
  - [18] Gabor N.M. et al. Hot Carrier-Assisted Intrinsic Photoresponse in Graphene. *Science* **334**, 648-652 (2011)
  - [19] Song, J.C.W. et al. Hot Carrier Transport and Photocurrent Response in Graphene. *Nano Lett.* **11**, 4688-4692 (2011)
  - [20] Kampfrath, T. et al. Strongly coupled optical phonons in the ultrafast dynamics of the electronic energy and current relaxation in graphite. *Phys. Rev. Lett.* **95**, 187403 (2005).
  - [21] Bistritzer, R. & MacDonald, A. H. Electronic cooling in graphene. *Phys. Rev. Lett.* **102**, 206410 (2009).
  - [22] Song, J.C.W., Reizer, M.Y. & Levitov, L.S. Disorder-Assisted Electron-Phonon Scattering and Cooling Pathways in Graphene. *Phys. Rev. Lett.* **109**, 106602 (2012).

- [23] Graham, M.W., Shi, S.-F., Ralph, D.C., Park, J. & McEuen, P.L. Photocurrent measurements of supercollision cooling in graphene. *Nature Phys.* **9**, 103-108 (2013)
- [24] Betz, A.C. et al. supercollision cooling in undoped graphene. *Nature Phys.* **9**, 109-112 (2013)
- [25] Graham, M.W. et al. Transient Absorption and Photocurrent Microscopy Show That Hot Electron Supercollisions Describe the Rate-Limiting Relaxation Step in Graphene. *Nano Lett.* **13** 54975502 (2013)
- [26] Low, T. et al. Cooling of photoexcited carriers in graphene by internal and substrate phonons. *Phys. Rev. B* **86**, 045413 (2012)
- [27] Hamm, J.M. et al. Nonequilibrium plasmon emission drives ultrafast carrier relaxation dynamics in photoexcited graphene. *Phys. Rev. B* **93**, 041408 (2016)
- [28] Principi, A. et al. Super-Planckian electron cooling in a van der Waals stack. *Arxiv* **1608.01516v1** (2016)
- [29] D. Brida et al., Ultrafast collinear scattering and carrier multiplication in graphene. *Nature Comm.* **4**, 1987 (2013)
- [30] Banszerus, L. et al. Ultrahigh-mobility graphene devices from chemical vapor deposition on reusable copper. *Science Adv.* **1**, e1500222 (2015)
- [31] Crossno, J. et al. Development of high frequency and wide bandwidth Johnson noise thermometry. *App. Phys. Lett.* **106**, 023121 (2015)
- [32] Nyquist, H. Thermal Agitation of Electric Charge in Conductors. *Phys. Rev.* **32**, 110113 (1928).
- [33] Tomadin, A, et al. Accessing phonon polaritons in hyperbolic crystals by angle-resolved photoemission spectroscopy. *Phys. Rev. Lett.* **115**, 087401 (2015)
- [34] Giuliani, G.F. & Vignale G. Quantum Theory of the Electron Liquid (Cambridge University Press, Cambridge, 2005)
- [35] Ma, Q. et al. Competing Channels for Hot-Electron Cooling in Graphene. *Phys. Rev. Lett.* **112**, 247401 (2014)
- [36] Jadidi, M.M. et al. Infrared Nonlinear Photomixing Spectroscopy of Graphene Thermal Relaxation. *Arxiv* **1607.02181** (2016)
- [37] Das Sarma, S., Adam, S., Hwang, E. H. & Rossi, E. Electronic transport in two-dimensional graphene. *Rev. Mod. Phys.* **83**, 407 (2011)
- [38] Hwang, E. H. & Das Sarma, S. Acoustic phonon scattering limited carrier mobility in two-dimensional extrinsic graphene. *Phys. Rev. B* **77**, 1 (2008)
- [39] Principi, A. et al. Plasmon losses due to electron-phonon scattering: The case of graphene encapsulated in hexagonal boron nitride. *Phys. Rev. B* **90**, 14 (2014)
- [40] Hwang, E.J., Rossi, E. & Das Sarma, S. Theory of thermopower in two-dimensional graphene. *Phys. Rev. B.* **80**, 235415 (2009)

## Appendix

*Extraction of the deformation potential from transport measurements* — Momentum-non-conserving collisions can occur because of: *i*) long-range scattering mechanisms, *ii*) short-range scattering mechanisms, and *iii*) electron-phonon scattering [37]. If momentum scattering is solely determined by electron-phonon interaction (process *iii*), we can use the measured graphene mobility  $\mu$  at a given carrier density  $n$  to obtain the deformation potential following Refs. [38, 39]:  $\mu = \frac{4\hbar v_F^2 e \rho v_s^2}{\pi D^2 n k_B T}$ , where  $v_F$  is the Fermi velocity,  $\rho$  the mass density,  $v_s$  the sound velocity and  $\hbar$ ,  $e$  and  $k_B$  the reduced Planck constant, electron charge and Boltzmann constant, respectively. Inserting relevant numbers gives  $D = 35$  eV. However, since the mean free path for low carrier concentrations scales linearly with  $n$  (see Fig. 2c), we know that long-range scattering (process *i*) also play a role [37] and therefore this value is an *upper* limit.

*Extraction of the deformation potential from cooling dynamics* — We follow Ref. [23] in calculating the experimental photovoltage dip  $\Delta V_{\text{PTE}}$ , based on photo-thermoelectric photovoltage generation in combination with supercollision cooling. Since this model requires that  $T_{\text{el}} \ll T_F$ , we apply this analysis to our data at  $n = 1.7 \times 10^{12}/\text{cm}^2$ , where  $T_F = \frac{E_F}{k_B} \approx 1800$  K. We numerically solve the energy dissipation rate  $C_n \frac{dT_e}{dt} = -A(T_e^3 - T_L^3)$  with the heat capacity (for non-neutral graphene)  $C_n = \alpha T_{\text{el}}$  and  $\frac{A}{\alpha} = 0.47 \frac{1}{k_F \ell} \frac{D^2}{\rho v_s^2} \frac{E_F}{(\hbar v_F)^2} \frac{k_B}{\hbar}$  to obtain the temperature dynamics  $T_e(t)$ . The photovoltage  $V_{\text{PTE}}$  at a delay time  $\Delta t$  between two pulses follows from  $V_{\text{PTE}} = \int_0^{\Delta t} V(t, T_1) dt + \int_{\Delta t}^{\infty} V(t - \Delta t, T_2) dt$ . Here we use the instantaneous photo-thermoelectric voltage  $V(t, T) = BT(T - T_L)$ , where  $B$  is a proportionality constant related to the Seebeck effect. The initial hot-electron temperature after the first (second) laser pulse  $T_1$  ( $T_2$ ) is given by  $\sqrt{T_{\text{in}}^2 + T_{\text{add}}^2}$ , with  $T_{\text{in}}$  the temperature before arrival of the laser pulse, and  $T_{\text{add}}$  the temperature equivalent to the added pulse energy. We obtain a photovoltage dip by repeating this calculation while varying  $\Delta t$ , and fit the data to extract the deformation potential, finding  $D = 65$  eV (see Fig. A5). We note that using Ref. [22] to relate  $\frac{A}{\alpha}$  to  $D$  would give a  $D$  that is a factor  $\sqrt{8}$  higher.

*Extraction of the initial hot carrier temperature* — The measured photovoltage  $V_{\text{PTE}}$  scales with the *time-averaged* increase in electron temperature  $\langle T_e - T_L \rangle$ . In the case of cooling length smaller than the laser spot size (which is the case at room temperature) the relevant heat equation gives a simple linear scaling between peak and average increase in electron temperature, governed by the heat conductance. We assume that the heat conductance and Seebeck coefficient are constant with

power [40], so that the photovoltage as a function of power directly represents the peak electron temperature increase as a function of power:  $V_{\text{PTE}} = a(T_e - T_L)$ . For undoped graphene, the peak electron temperature  $T_e$  after illumination with laser power  $P$  is  $T_e = \sqrt[3]{T_L^3 + bP}$ , where  $b$  is a constant that depends on laser repetition rate  $f$ , absorption coefficient  $\eta_{\text{abs}}$ , heating efficiency  $\eta_{\text{heat}}$  and spot size  $L_{\text{spot}}$ . The cube root comes from the  $T^2$ -scaling of the electronic heat capacity for neutral graphene (we apply this analysis to a low carrier density of  $0.06 \times 10^{12}/\text{cm}^2$ ). By fitting  $V_{\text{PTE}}$  as a function of  $P$ , we extract the constants  $a$  and  $b$ , which allows us to recover the power-dependent initial carrier temperature  $T_e$ . We verify the obtained constant  $b$ , and find good agreement when using an absorption of  $\eta_{\text{abs}} = 1\%$  (due to the layered dielectrics), a heating efficiency of  $\eta_{\text{heat}} = 80\%$  and a spot size of  $L_{\text{spot}} \sim 2 \mu\text{m}$ .

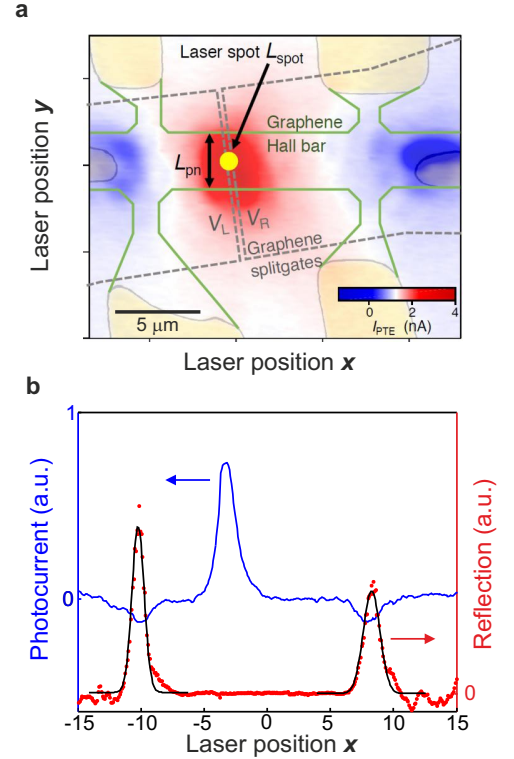


FIG. A1: **Photocurrent and reflection data.** **a)** Spatial image, where the laser focus is scanned over the device, while photocurrent and laser reflection are simultaneously measured (at room temperature). The reflection image indicates the metal contacts (yellow) and we observe negative photocurrent at the metal-graphene interfaces and positive photocurrent at the *pn*-junction. **b)** Photovoltage and spatial-derivative reflection line traces as indicated in panel **b**, with Gaussian fits to the reflection image to determine the spot size. The spatial extent of the photocurrent is similar to the spot size for these room temperature measurements. This shows that the cooling length is relatively short, indicating that on a micron length-scale lateral heat transport is still slower than hot-carrier cooling (at room temperature).



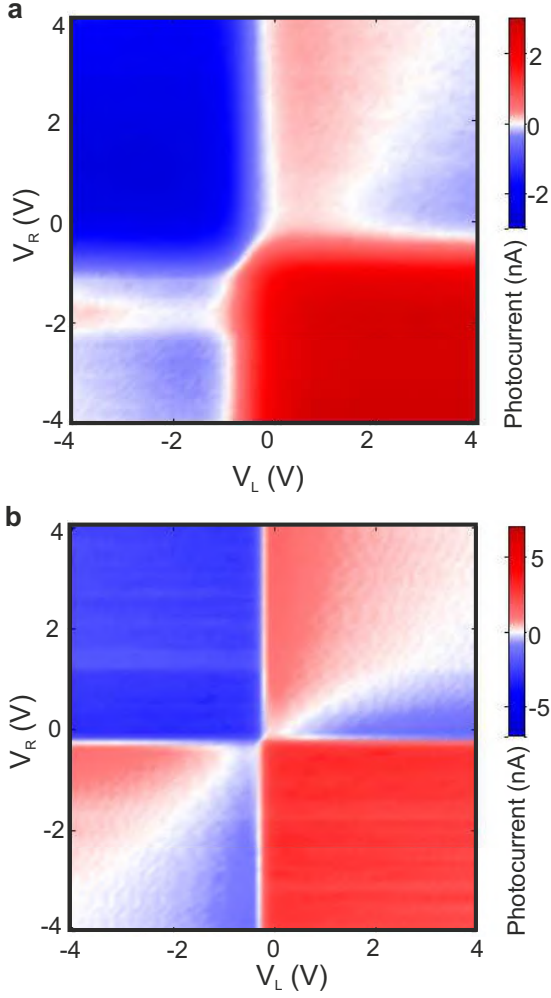


FIG. A2: **Characteristic sixfold pattern for photothermoelectric effect.** The photovoltage at the interface between the two graphene regions, for single-pulse excitation as a function of voltage on the two graphene split gates  $V_L$  and  $V_R$  at room temperature (a) and 30 K (b). The sixfold patterns indicate photo-thermoelectric photovoltage generation, which scales with the light-induced increase in carrier temperature [18, 19].

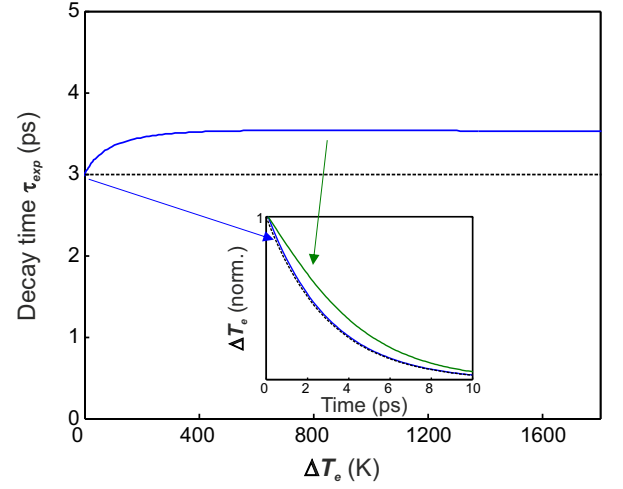
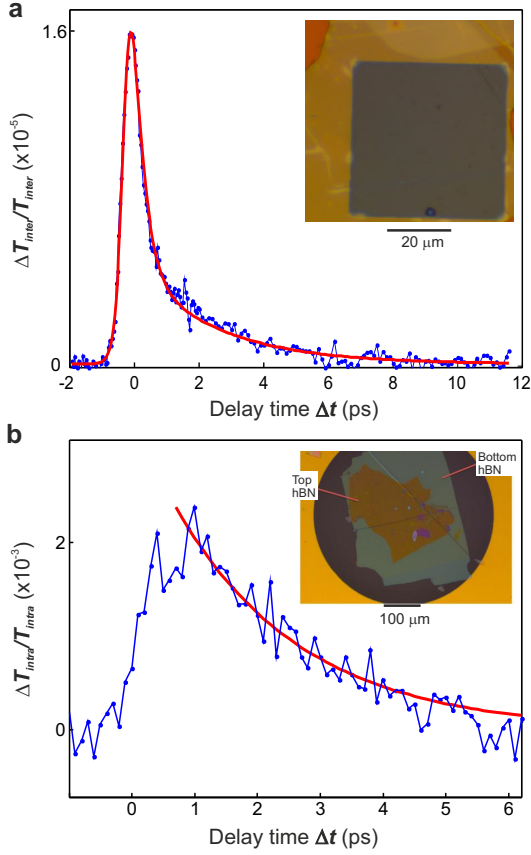
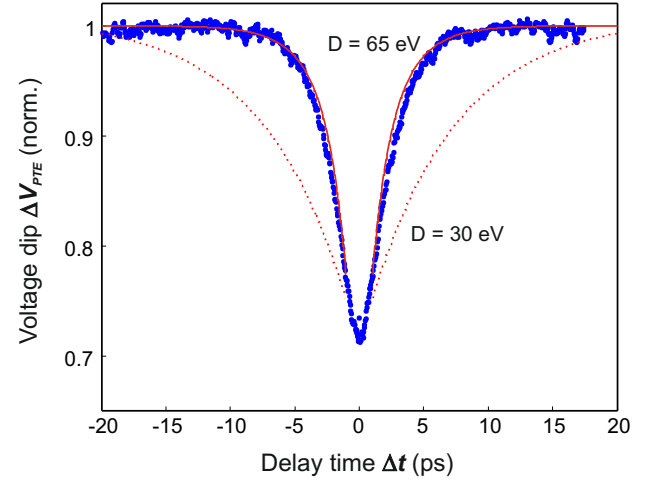


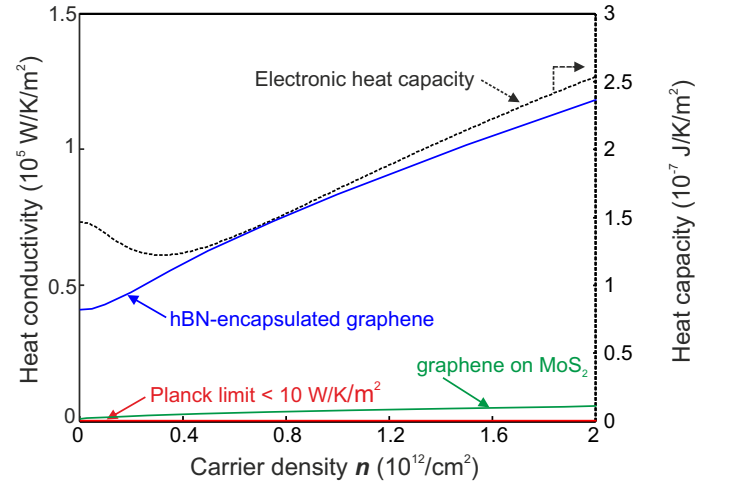
FIG. A3: **Effect of nonlinearity on measured dynamics.** To study the relation between the observed dynamics of the photovoltage dip and the actual cooling dynamics, we calculate the photovoltage dip dynamics for a given exponential cooling time of 3 ps. To this purpose, we integrate the generated photocurrent as a function of real time, after an ultrafast laser pulse heats up the carriers to a certain temperature that is related to the incident power and a second pulse heats up the carriers to a temperature that depends on the residual heat in the carrier system (see Methods). The higher the power, the higher the peak temperature and the stronger the nonlinearity (due to the temperature-dependent electronic heat capacity). We find that the photocurrent dip as a function of delay time between the two ultrafast pulses gives an overestimation of the cooling dynamics by  $<20\%$ . The inset shows the 'real' cooling dynamics (black dashed line), the photovoltage dynamics for very low power (blue line) and for a laser power that corresponds to  $T_e \approx 1000$  K (green line).



**FIG. A4: Verification of RT decay time using pump-probe spectroscopies.** **a)** Ultrafast optical pump – optical probe spectroscopy on hBN-encapsulated CVD graphene, prepared as in Ref. [30] and fully covering a  $50\ \mu\text{m}$  square aperture on a transparent  $\text{SiO}_2$  substrate (see inset). The ultrashort pump pulses, with a wavelength of  $780\ \text{nm}$ ,  $100\ \text{fs}$ -duration, and incident fluence of  $\sim 10\ \mu\text{J}/\text{cm}^2$ , creates a non-equilibrium distribution that quickly thermalizes through electron-electron scattering [29]. This modified, hot-carrier distribution affects available interband transitions through Pauli blocking, which we probe using ultrashort pulses at  $1.3\ \mu\text{m}$ . The differential transmission  $\Delta T_{\text{inter}}/T_{\text{inter}}$  presents a bi-exponential decay, with the slower component representing the main hot-electron cooling channel with a time scale of  $\sim 2.5\ \text{ps}$ . The ultrafast initial decay, with  $300\ \text{fs}$  time constant, is due to cooling to strongly coupled optical phonons [20], in particular of carriers with energy above the optical phonon energy of  $\sim 0.2\ \text{eV}$ . This optical phonon cooling thus occurs at very high  $T_e$  and dominates only during the first few hundred femtoseconds of carrier cooling. This decay is not seen in time-resolved photocurrent and optical pump – terahertz probe measurements since these techniques both have a sub-linear temperature response (see also Fig. A7) and therefore are not sensitive to this initial cooling process. **b)** The pump-induced change in transmitted terahertz (THz) signal as a function of pump-probe delay time  $\Delta t$  using an hBN-encapsulated CVD graphene sandwich with dimensions  $> 200\ \mu\text{m}$ , prepared as in Ref. [30] and transferred inside a  $500\ \mu\text{m}$  round aperture (see inset). The loosely focused pump light, with a wavelength of  $800\ \text{nm}$  and pulse energy density of  $\sim 20\ \mu\text{J}/\text{cm}^2$  creates hot carriers and a modified carrier distribution. Terahertz light with frequency  $0.4 - 1.2\ \text{THz}$  subsequently probes the effect of the modified distribution on intraband transitions, i.e. the electronic response. The pump-probe dynamics of the differential THz transmission  $\Delta T_{\text{intra}}/T_{\text{intra}}$  represent hot-carrier cooling (for carriers with energy below the optical phonon energy) and give a timescale of  $\sim 2.2\ \text{ps}$ .



**FIG. A5: Comparing data with supercollision model.** The photovoltage dip dynamics ( $T_L = 300\ \text{K}$  and  $n = 1.7 \times 10^{12}/\text{cm}^2$ , compared with the dynamics according to the supercollision cooling model with a deformation potential of  $65\ \text{eV}$  (solid red line) and  $30\ \text{eV}$  (dashed red line). The model takes into account that we are in the strong heating regime with  $T_e = 1500\ \text{K}$  (see Methods).



**FIG. A6: Interfacial heat conductivity and heat capacity.** Since the hyperbolic cooling time  $\tau_{\text{calc}}^*$  is determined by the ratio of the interfacial heat conductivity  $\Gamma$  and the electronic heat capacity  $C_n$ , we plot these separate entities as a function of carrier density  $n$ . We also show the interfacial heat conductivity for  $\text{MoS}_2$  hyperbolic phonons, and for black-body radiation of graphene in vacuum, which is orders of magnitude lower than for hBN or  $\text{MoS}_2$ .

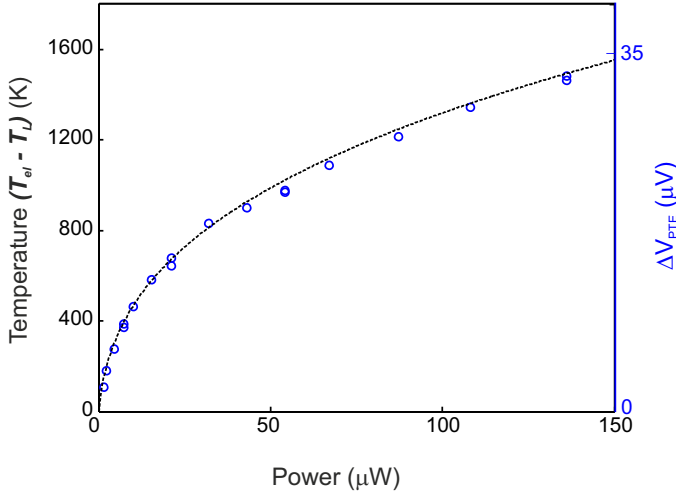


FIG. A7: **Estimation of hot carrier temperature.** The photovoltage as a function of laser power (right vertical axis) and the extracted laser-induced temperature increase (left vertical axis) for undoped graphene, using the analysis described in the Methods section.

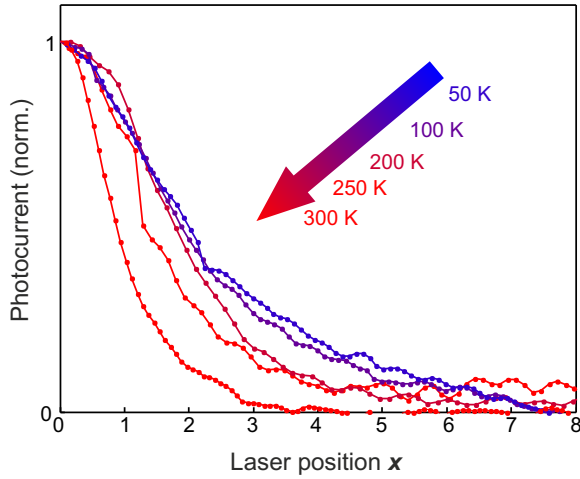


FIG. A8: **Spatial extent of photocurrent.** Spatial line traces of photocurrent vs. laser position for five different lattice temperatures. Position 0 corresponds to the  $pn$ -junction. At lower temperatures, the spatial extent of the photocurrent increases. This is caused by increased mechanical vibrations of the sample (due to the compressor of our helium closed cycle cryostat) and by a longer cooling length (due to slower hyperbolic hot-carrier cooling and a longer mean free path). This means that lateral heat diffusion out of the laser spot leads to an additional cooling channel, i.e. Wiedemann-Franz cooling. As a result of the larger photocurrent spot size the initial temperature is significantly lower for lower lattice temperature.

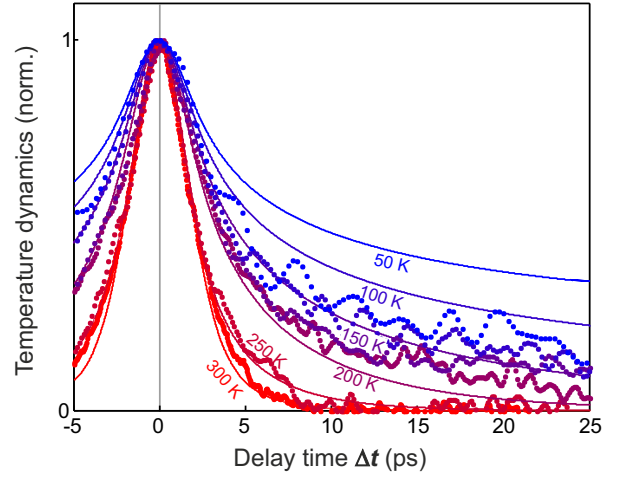


FIG. A9: **Cooling dynamics at high doping.** Comparison of the complete cooling dynamics as measured (data points) and calculated (solid line) for graphene with  $n = 1.7$  and  $2 \times 10^{12}/\text{cm}^2$ , respectively, and varying lattice temperature. We use an initial temperature increase of  $\Delta T = 300$ – $500$  K for all temperatures, except  $1250$  K for  $T_L = 300$  K, the same as we did for undoped graphene. We see deviations between data and model starting from  $T_L = 150$  K, most likely caused by normal collision with graphene acoustic phonons [21] becoming the dominant cooling mechanism, as also observed in Ref. [35].